ORNL/TM-13243

Computer Science and Mathematics Division

Mathematical Sciences Section

BETA TESTING THE CHEN-1000 SERVER

Thomas H. Dunigan

Mathematical Sciences Section
Oak Ridge National Laboratory
P.O. Box 2008, Bldg. 6012
Oak Ridge, TN 37831-6367
thd@ornl.gov

Date Published: June 1996

Research was supported by the Applied Mathematical Sciences Research Program of the Office of Energy Research, U.S. Department of Energy.

Prepared by the
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831
managed by
Lockheed Martin Energy Research Corp.
for the
U.S. DEPARTMENT OF ENERGY
under Contract No. DE-AC05-96OR22464

Contents

BETA TESTING THE CHEN-1000 SERVER

Thomas H. Dunigan

Abstract

This report summarizes the beta testing of Chen System's eight processor CHEN-1000 server at Oak Ridge National Laboratory in the fall of 1995. The performance of the shared-memory multiprocessor is measured and compared with other shared-memory multiprocessors.

1. Introduction

The Department of Energy's Oak Ridge National Laboratory (ORNL) has been conducting research in high performance scientific computing for several decades. Since the mid 1980's, ORNL has been a leader in the evaluation and use of massively parallel computers. ORNL has developed applications and mathematical algorithms to exploit parallelism, and in the 1990's, DOE selected ORNL as one of its high performance computing centers as part of the government's High Performance Computing and Communications (HPCC) initiative.

One component of ORNL's research in computational science is the evaluation of "early systems." ORNL has worked with computer vendors in beta testing early models of various parallel processors. These evaluations involve benchmarking the systems, porting and customizing applications, and evaluating the effectives of innovative computer architectures in solving scientific applications. In 1986, ORNL evaluated the first Intel iPSC/1 [2], then beta tested the later generation iPSC/2 [3] and iPSC/860 [4], Paragon [6], and Paragon MP [8]. In addition to these distributed-memory parallel processors, ORNL evaluated the first Kendall Square distributed shared-memory multiprocessor [5], [7].

In 1995, Chen Systems loaned ORNL one of their CHEN-1000 multiprocessors for beta testing. Although aimed at the commercial server market, the CHEN-1000 was compared with other multiprocessors that ORNL has evaluated for scientific computations. This report details the performance of the shared-memory CHEN-1000.

In the following section, the CHEN-1000 architecture is summarized. Section 3 describes the performance of the shared-memory architecture and compares its performance to other shared-memory and distributed-memory architectures. Appendix A provides a time-line of the events during the test and evaluation of the CHEN-1000 system, and Appendix B summarizes the architectures of other multiprocessors that were compared with the CHEN-1000.

2. CHEN-1000 Architecture

The Chen Systems CHEN-1000 server is shared-memory multiprocessor based on the Intel Pentium processor. Symmetric multiprocessing is controlled by the UnixWare 2 operating system. The server provides high I/O throughput via two PCI buses that can double data throughput when configured as a dual PCI disk subsystem. The server can support almost two terabytes of on-line storage, and the High Throughput File System (HTFS) provides transaction-oriented caching and journaling.

The system tested at ORNL during 1995 consisted of 8 100 MHz Pentiums sharing 512 MB of memory over a 400 MB/second bus. Each processor had an 8KB L1 cache and a 512KB L2 cache. The system included 40 GB of disk space, system console, and diagnostic console. A C compiler was provided, and later ORACLE and a bus monitor were installed on the system as well.

3. Performance

Our first test of the Chen system was to examine the stability and usability of UnixWare. We compiled and ran several traditional UNIX benchmark codes (BYTE benchmark, fsx, hint, ttcp) and various software packages (ntp, GNU utilities, paranoia). We also built some security extensions (tcpwrappers, logdaemon, skey, tripwire). A few of the auto-build facilities failed to identify UnixWare, but manual System V building usually succeeded.

UnixWare is a stable version of UNIX, and its multiprocessor support appears quite solid. All of the utilities that we tested performed to specifications, except for ntp. We were not able to get tight clock synchronization with ntp. UnixWare appeared to be resetting the clock every 30 minutes to an internal time-of-day clock.

We measured the floating point performance of the 100 MHz Pentium at 10.7 Megaflops on a C LINPACK-100 [1]. Many of our other computational benchmarks and applications are written in FORTRAN. FORTRAN was not provided with our beta system, so we did not do as extensive testing as we have on other multiprocessors. In addition, we have several parallel FORTRAN applications that we use for performance analysis, those tests were omitted as well.

We compared thread and fork creation, lock and unlock, barriers, and concurrent update of shared variable (no locks). Table 3.1 compares single processor performance of the 100 MHz Chen Penitum with single processors on the Intel Paragon MP, KSR, BBN, and Sequent. The Intel Paragon uses a 50 MHz i860 processor. (It should be noted that even Intel considered the Intel Pentium unsuitable for scientific computation and chose to use the super-scalar i860 processor as the basis of its iPSC/860 and Paragon series. The next generation Intel multiprocessor will use the Pentium Pro, however). The KSR is a ring-based shared-memory multiprocessor using a 20 MHz custom processor. The Sequent Symmetry is bus-based shared-memory multiprocessor using 16 MHz 386 processors. The BBN TC2000 is a cascaded-switch based shared-memory multiprocessor using 20 MHz M88000 processors (also see Appendix B).

Time on one CPU (μs)					
	Chen	MP	KSR	BBN	Sequent
fork/wait	2,933	50,000	108,000	44,000	14,000
thread/join	548	1,191	130	79	26
lock/unlock	1	13	3	8	10
barrier	1	13	39	11	10
hotspot	0.03	0.12	0.41	1.3	1.4

Table 3.1: *Single processor performance of multiprocessing primitives.*

Since our test tools are aimed at scientific calculations, we could not properly evaluate Chen's high-performance file system. We did run simple I/O and file benchmarks and found no anomalous behavior, and performance exceeded that of our local workstations. We ran TCP and UDP throughput and latency tests over Ethernet and found the CHEN-1000 provided near full-bandwidth performance, comparable to most high-end workstations. Our IP multicast tests failed because multicast was not supported by UnixWare.

Our other tests evaluated parallel performance of the CHEN-1000. We ported a message-passing simulator that utilizes System V shared-memory and *fork()* for coarse-grained parallelism. This simulator ran successfully and gave linear speedups on some simple application kernels. Most of our parallel applications are based on a finer grain parallelism through FORTRAN parallel DO-loops or threads, so we also evaluated the thread management on the CHEN-1000.

Our most recent performance analysis has been on the Paragon MP system, which has three processors sharing memory on a node, so we compare three-processor performance with the CHEN-1000. Table 3.2 compares the performance of three processors doing a lock/unlock, barrier, or accessing the same memory location. There appear to be no architecture or implementation penalties in the thread primitives for one, two, or three processors.

Time on three CPUs (μs)					
	Chen	MP	KSR	BBN	Sequent
lock/unlock	3	104	17	20	38
barrier	93	110	119	38	24
hotspot	.03	0.81	2.5	5.3	6.4

Table 3.2: *Three processor performance.*

Figures 3.1 and 3.2 compare lock and barrier performance of the CHEN-1000 to other shared and distributed memory multiprocessors. The lock performance of the CHEN-1000 compares quite favorably to the older multiprocessor systems.

The spin barrier (*_barrier_spin()*) was used in the Chen barrier evaluations. The spin barrier is equivalent to the barrier implementations found on the comparison multiprocessors. The CHEN-1000 barrier times are somewhat worse than the other shared memory multiprocessors. The Delta, iPSC/860, and Paragon have slower barriers, but these are distributed-memory multiprocessors and scale to thousands of processors. Bus-based shared-memory multiprocessors are typically limited to twenty or thirty processors. These performance results indicate the CHEN-1000 scales well through eight processors.

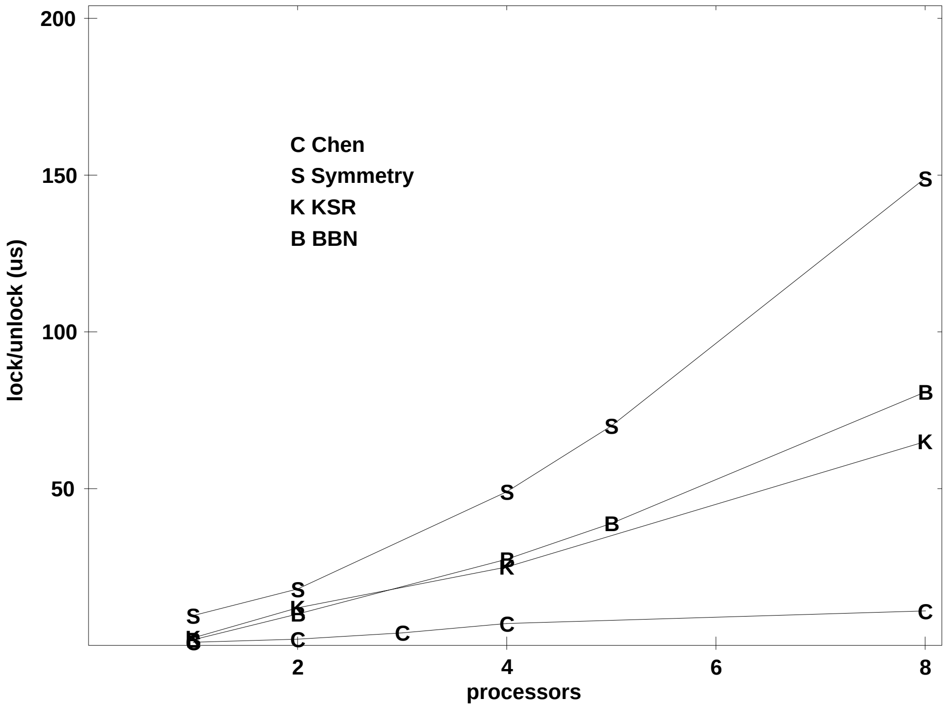

Figure 3.1: *Average lock/unlock time for a single lock.*

Table 3.3 compares the CHEN-1000 performance on three processors over a set of application kernels in C. The same copy of the code was run on each multiprocessor, and the codes have not been tuned. The numeric integration kernels effectively operate from cache, so near linear speed-up is achieved. The Cholesky code is a little more memory intensive, and the slower performance results from lock and bus contention.

Figure 3.3 shows the C double-precision, floating point performance for a Cholesky factorization of a 400×400 matrix. The 0 processor megaflop rating is the speed of the serial code, as opposed to the parallel code run on one processor. Again, the CHEN-1000 shows near linear speed-up.

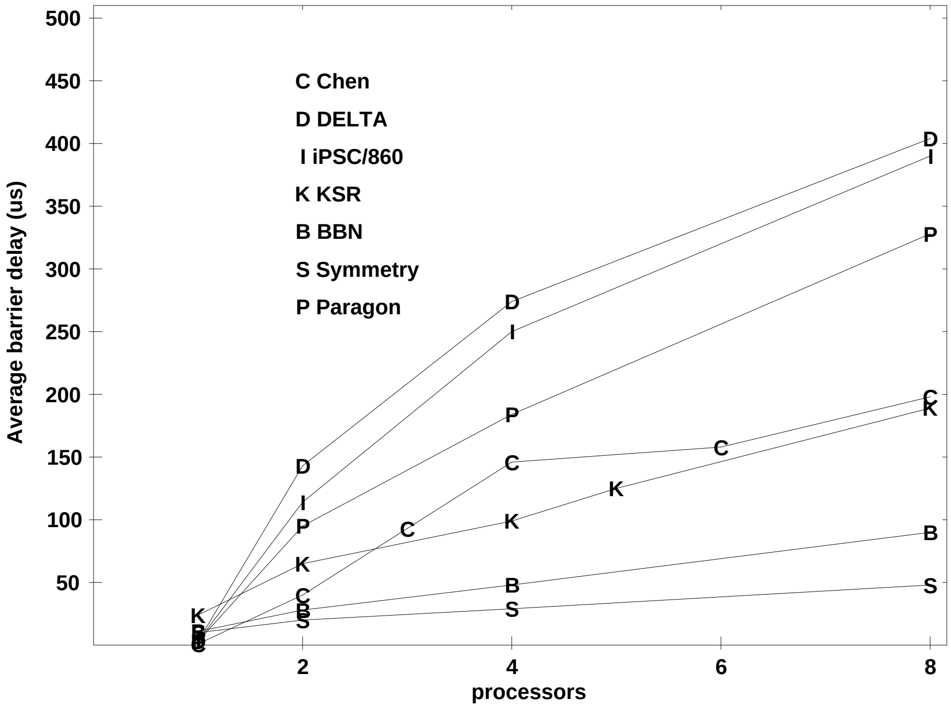

Figure 3.2: *Average time for a barrier synchronization.*

Speedup on three CPUs				
	Chen	MP	KSR	Sequent
Integration (C)	2.9	2.9	2.9	3.0
Jacobi iteration (C)	2.0	2.7	2.8	3.0
Cholesky (1K × 1k) (C)	2.0	2.1	3.0	2.9

Table 3.3: *Speedup on three CPUs for various application kernels.*

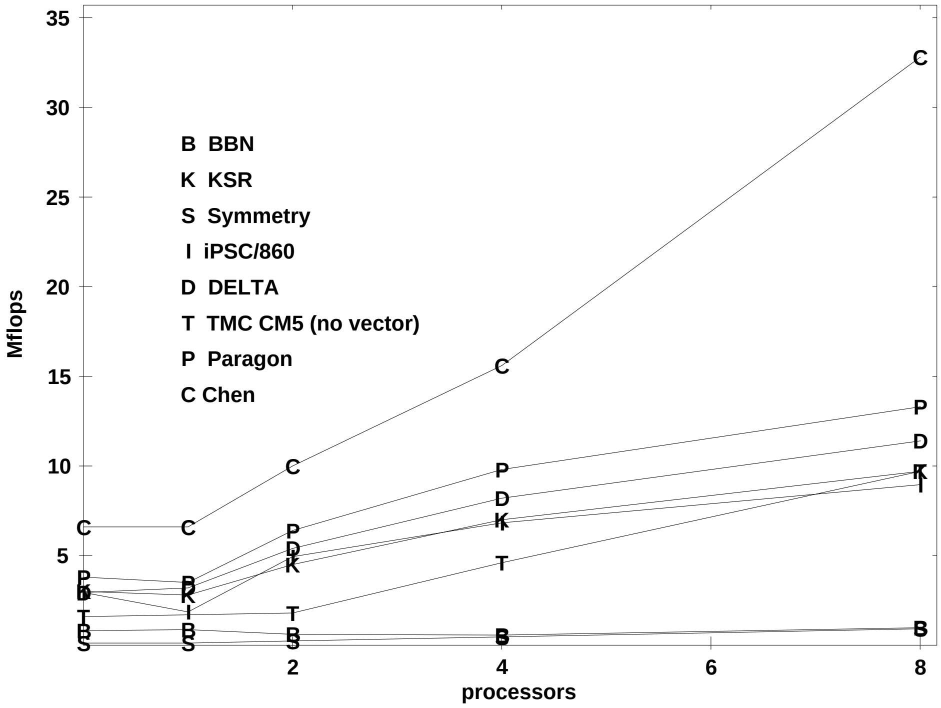

Figure 3.3: *Megaflops for* 400×400 *double-precision C Cholesky.*

4. Summary

Despite being a beta system, the CHEN-1000 system proved to be a reliable multiprocessor system with adequate multiprocessor support from UNIXWare. The shared-memory performance of the CHEN-1000 is competitive with other older shared-memory multiprocessors. The I/O performance was good, and hardware and software performance monitoring should be effective in tuning parallel applications. The system is best suited for coarse grain parallelism, and no automatic parallelization was provided by the compilers in our beta unit.

5. References

[1] J. Dongarra. Performance of various computers using standard linear equations software in a Fortran environment. Technical Report CS-89-85, University of Tennessee, 1995.

[2] T. H. Dunigan. Performance of three hypercubes. Technical report, Oak Ridge National Laboratory, Oak Ridge, TN, 1987. ORNL/TM-10400.

[3] T. H. Dunigan. Performance of a second generation hypercube. Technical report, Oak Ridge National Laboratory, Oak Ridge, TN, 1988. ORNL/TM-10881.

[4] T. H. Dunigan. Performance of the Intel iPSC/860 and Ncube 6400 hypercubes. *Parallel Computing*, 17:1285 – 1302, 1991.

[5] T. H. Dunigan. Kendall Square multiprocessor: Early experiences and peformance. Technical report, Oak Ridge National Laboratory, 1992. ORNL/TM-12065.

[6] T. H. Dunigan. Early experiences and performance of the Intel Paragon. Technical report, Oak Ridge National Laboratory, 1993. ORNL/TM-12194.

[7] T. H. Dunigan. Multi-ring performance of the Kendall Square multiprocessor. Technical report, Oak Ridge National Laboratory, 1993. ORNL/TM-12331.

[8] T. H. Dunigan. Beta testing the intel paragon mp. Technical report, Oak Ridge National Laboratory, 1995. ORNL/TM-12830.

[9] R. P. LaRowe and C. S. Ellis. Experimental comparison of memory management policies for numa multiprocessors. Technical report, Duke University, April 1990. CS-1990-10.

[10] R. D. Rettberg, W. R. Crowther, P. P. Carvey, and R. S. Tomlinson. The monarch parallel processor hardware design. *Computer*, 23:18–30, April 1990.

Appendix

A. Timeline

July, 1995. 4-CPU Chen system installed.

August, 1995. UNIX benchmarks and initial parallel testing.

September, 1995. Security and network testing.

October, 1995. Upgrade to 8 CPUs, 512M, and bus monitor.

November, 1995. Complete parallel benchmarks and I/O testing.

December, 1995. Data base and application testing.

B. Comparative Architectures.

The report compares shared memory performance of the CHEN-1000 with several shared-memory multiprocessors. A summary of the other shared-memory multiprocessors compared with the CHEN follows.

BBN TC2000

The BBN TC2000 at Argonne National Laboratory (ANL) is a 45 processor shared-memory parallel processor. Each processor is a Motorola 88000 running at 20MHz with 16 MB of memory fronted by a 16KB data cache and a 16KB instruction cache. All of the memories are interconnected by a 2-stage 8-way switch. The system can be expanded up to 512 processors. The Uniform programming environment (under nX 2.0.6) provides the program with both local and explicitly allocated shared memory. The shared memory may be allocated in another processor's memory, and thus a non-uniform memory access (NUMA) model is supported. In the absence of contention, a remote reference typically takes less than two microseconds, and a single channel of the switch has a bandwidth of 40 MB/second [10]. The architecture could be used with other memory management policies [9]. Compiles on the BBN were done with -O -lus. LINPACK 100×100 double-precision on a single processor was 1.0 Mflops using -OLM -autoinline.

Intel Delta

The Intel Delta is a 512-node mesh designed as a prototype for the Intel Paragon family. Each node has a 40 MHz i860 with 8 KB cache and 16 MB of memory. Communication channels have a peak rate of 22 MB/second. LINPACK 100×100 double-precision on a single processor using -Minline=daxpy -Knoieee was 9.8 Mflops. Tests were performed using NX 3.3.10.

Intel iPSC/860

The Intel iPSC/860 is Intel's third generation hypercube. Each node has a 40 MHz i860 with 8 KB cache and at least 8 MB of memory. Communication channels have a peak rate of 2.8 MB/second. LINPACK 100×100 double-precision on a single processor was 9.8 Mflops using -Minline=daxpy -Knoieee. Tests were performed using NX 3.3.2.

Intel Paragon

The Intel Paragon is a distributed-memory multiprocessor with shared memory on each node. A node on the Paragon MP supports a shared-memory architecture with three 50 MHz i860XP processors sharing a 400 MB/second bus and memory. Each i860XP has its own 16 KB data and instruction cache, and each node has at least 64 MB of memory. The 50 MHz i860XP is a super-scalar architecture capable of a peak 75 Mflops (double precision). LINPACK 100×100 double-precision on a single processor was 11 Mflops ([1]).

Kendall Square

The Kendall Square uses custom-designed 20 MHz processors that share memory on a one gigabyte per second ring. Each processor has a 256KB cache, and the global memory is managed as a cache. A single processor generates a maximum of 40 MB/second against the ring. LINPACK 100×100 double-precision on a single processor was 15 Mflops [5].

Sequent Symmetry

The 26 processor Sequent Symmetry located at ANL is based on 80386/387 processors (16 MHz) with a Weitek 3167 floating point co-processor. Each processor has a 64KB cache, and 32 MB of memory is shared by all processors on a 54 MB/second bus. The maximum configuration is 30 processors. The processors run Dynix 3.1.2, and compiles were done using -O. LINPACK 100×100 double-precision on a single processor was 0.37 Mflops [1].

TMC CM5

The CM5 is a hypertree multiprocessor using 32 MHz SPARC processors with four vector units and 16 MB of memory per node. LINPACK 100×100 double-precision on a single processor using the vector units was 68 Mflops [1]. Communication channels have a peak rate of 20 MB/second.